Another way to Calculate the Standard Corrected Time of a Yacht Race

R.E. Deakin
Dunsborough, WA, 6281, Australia
email: randm.deakin@gmail.com

24-Dec-2024

Abstract

In yacht racing, a Performance Handicap System yields a handicap (a number) that enables yachts of varying speed potential to compete in races where the yacht with the least corrected time wins – where corrected time is elapsed time multiplied by the handicap. After a race, calculated handicaps are obtained from a set of rules applied after the Standard Corrected Time *SCT* has been determined. This paper demonstrates a new method of determining the *SCT* and reviews current methods.

Introduction

A Performance Handicap System (PHS) in yachting is a set of rules and mathematical calculations that enable yachts of varying speed potential to compete in races where the yacht with the least corrected time is the winner. The PHS produces a *handicap* which is a number, usually somewhere between 0.750 and 1.250. In local terminology (Australia) this handicap is known as the *Allocated Handicap AHC* and the yacht's *Elapsed Time ET* multiplied by the allocated handicap yields the *Corrected Time CT,* or

$$CT_k = ET_k \times AHC_k \tag{1}$$

where the subscript k denotes the k^{th} boat in the fleet of n yachts and $k = 1, 2, 3, \dots, n$ and

$$ET_k = \text{ finish time of } k^{th} \text{ yacht} - \text{start time} \tag{2}$$

So, the handicap is a numerical value that measures the performance of both the yacht and the crew.

[A yacht's PHS-derived handicap is different from a yacht's *rating* which is a numerical measure of potential speed based upon the yacht's parameters, e.g., waterline length, beam, displacement, sail area, etc. and a sequence of mathematical formulas related to the physics of hydrodynamics and aerodynamics as applied to yachting force models (*World Sailing*[1] 2019). Several measurement systems give yacht ratings, e.g., The International Offshore Rule (IOR), the Chanel Rating System (CHS), the International Measurement System (IMS) and the International Rating Certificate (IRC). A popular measurement rating system in Australia, particularly in Victoria, is the Australian Measurement System (AMS) administered by Yacht Racing Services Association Inc. (YRSA) for the Australian yachting community. This paper is not concerned with measurement systems, or the ratings derived from them.]

The essence of a Performance Handicap System is the rules that enable the adjustment of handicaps after racing. Some of these rules may be arbitrary, some could be based on experience, and some could be in place to achieve desired outcomes. Indeed, a PHS used in one yacht club could be different from that used in another club; or the PHS used for a regatta could be different from the usual club PHS. And, as we will discuss, a *calculated handicap* produced by the PHS after a race is related to the number of yachts in that race; their handicaps; the handicap of a mythical or real *standard boat;* and the allowable changes in handicaps.

The sequence of steps in calculating a yacht's new PHS handicap are

1. A *Standard Corrected Time STC* is established for the race, and this is the corrected time of the *standard boat* (which may be real or mythical).

---

[1] *World Sailing* is the world governing body for the sport of sailing formed in 1907 and then known as the International Yacht Racing Union (IYRU). The name was changed to the International Sailing Federation (ISAF) in 1996 before adopting the name *World Sailing* in 2015.

2 A *Back Calculated Handicap BCH* is derived from the rule $ET_k \times BCH_k = SCT$ and

$$BCH_k = \frac{SCT}{ET_k} \tag{3}$$

There may be some screening of the *BCH* to detect anomalous results.

3 A *Performance Indicator PI* is calculated where

$$PI_k = BCH_k - AHC_k \tag{4}$$

4 A *Calculated Handicap CHC* is evaluated from a function of the allocated handicap and the performance indicator. There may be a further screening of the calculated handicap before it becomes the yacht's allocated handicap for the next race.

This paper will outline some of the usual methods of determining the standard corrected time and a new method called *Optimum Boat*.

Nomenclature

The following notation has been used

Symbol	**Meaning**	**Definition**
AHC	allocated handicap	
BCH	back-calculated handicap	$BCH = SCT/ET$
b	scale factor	$b \approx 1.4826$
c	tuning constant	$c = 4.685$
γ	(gamma) weighting constant	$\gamma = cS = 6.946(\mathrm{MAD})$
CHC	calculated handicap	
CT	corrected time	$CT = ET \times AHC$
ε	(epsilon) a small value	$\varepsilon = 0.001$
ET	elapsed time	$ET =$ finish time $-$ start time
ET_{AVE}	average of elapsed times	
$iter$	integer iteration counter	
k	integer counter	
M	median	
MAD	Median Absolute Deviation	
n	number of yachts in the race	
φ	(phi) objective function	$\varphi = \sum_{k=1}^{n} \rho(v_k)$
$\psi(v)$	(psi) influence function	$\psi(v) = \frac{d}{dv}\rho(v) = v\,w(v)$
PHS	performance handicap system	
PI	performance indicator	$PI = BCH - AHC$
q	ratio of times	$q = ET/ET_{AVE}$
$\rho(v)$	(rho) the arbitrary function of residuals	
σ	(sigma) population standard deviation	
$\hat{\sigma}$	estimate of population st. deviation	$\hat{\sigma} = b \times \mathrm{MAD}$
s	sample standard deviation	
σ^2, s^2	population and sample variance	
S	scale, a measure of the standard deviation of residuals	$S = b \times \mathrm{MAD} \approx 1.4826(\mathrm{MAD})$
SCT	standard corrected time	
v	residual, a small correction	$v = \hat{y} - y$

Symbol	Meaning	Definition
w	weight	$0 \le w \le 1$
$w(v)$	weight function	$w(v) = \psi(v)/v$
$y,\ \hat{y}$	a quantity y and its estimate $\hat{y}$	

Usual methods of determining the Standard Corrected Time *SCT* for a race

The Standard Corrected Time *SCT* for a race is the corrected time of the *standard boat,* sometimes called the *mark boat,* which may be a yacht in the fleet or an imaginary yacht. It is obtained from the list of corrected times of the yachts in the race where this list has been sorted in ascending order from least to greatest.

We will review three methods; ***Trimmed Fleet Average, 45% Boat*** and ***Median Boat*** as representative of the usual methods of determining the *SCT* of the race and use the Example Yacht Race results in Table 1 to calculate *SCTs* for each method.

The Example Yacht Race

The results of the Example Yacht Race shown in Table 1 have been taken from *World Sailing's*[2] *International Empirical Handicap Scheme for Yachts* shown in Appendix A. [The small error in the corrected time order of yachts C = *Charlie* and H = *Hotel* has been fixed in Table 1]

Sail No.	Yacht	Elapsed Time *ET*	Allocated Handicap *AHC*	Corrected Time *CT* (sec)	Place
10	*Juliet*	1:23:17 (4997 sec)	1.074	5366.778	**1**
7	*Golf*	1:32:29 (5549)	1.003	5565.647	**2**
1	*Alfa*	1:26:37 (5197)	1.079	5607.563	**3**
4	*Delta*	1:33:59 (5639)	1.008	5684.112	**4**
5	*Echo*	1:34:21 (5661)	1.005	5689.3050	**5**
6	*Foxtrot*	1:34:44 (5684)	1.004	5706.736	**6**
9	*India*	1:37:14 (5834)	0.982	5728.988	**7**
2	*Bravo*	1:42:36 (6156)	0.957	5891.292	**8**
8	*Hotel*	1:45:44 (6344)	0.948	6014.112	**9**
3	*Charlie*	1:48:24 (6504)	0.929	6042.216	**10**

Table 1. The Example Yacht Race times. Yachts are shown in Corrected Time order. Elapsed Time in hour, minute, second format as h:mm:ss and also in seconds (sec)

Trimmed Fleet Average

World Sailing in their *International Empirical Handicap Scheme for Yachts* (see Appendix A) use the average *CT* excluding the lowest 20% and highest 40% of the *CTs* (rounded down to whole numbers). That is, if there are 10 yachts in the race then $20/100 \times 10 = 2$ and $40/100 \times 10 = 4$, then the *CTs* of the first two, and the last four yachts are ignored and the *SCT* is the average of the *CTs* of yachts placed 3rd, 4th, 5th, and 6th. If there were 19 yachts in the race then $20/100 \times 19 = 3.8 \rightarrow 3$ and $40/100 \times 19 = 7.6 \rightarrow 7$, then the *CTs* of the first three, and the last seven yachts are ignored and the *SCT* is the average of the *CTs* of yachts placed from 4th to 12th.

[2] *World Sailing* is the governing body for the sport of sailing formed in 1907 and then known as the International Yacht Racing Union (IYRU). The name was changed to the International Sailing Federation (ISAF) in 1996 before adopting the name *World Sailing* in 2015.

In the Example Yacht Race shown in Table 1 where there are 10 yachts, the corrected times of *Juliet* and *Golf* (the lowest 20% of *CTs*) and *India, Bravo, Hotel* and *Charlie* (the highest 40% of *CTs*) are ignored and the *SCT* is the average of the *CTs* of *Alfa, Delta, Echo,* and *Foxtrot,* equal to 5672 sec = 1:34:32 (nearest second).

45% Boat

TopYacht[3] in their sailing software documentation suggest that from their experience, the *SCT* for the race be the *CT* of the "45% boat" where the 45% boat is the yacht finishing in 45th place on corrected time in a fleet of 100 yachts. If the 45% boat is not an integer (a whole number) then the nearest yacht is selected, e.g., in a fleet of 10, the 45% boat is $45/100 \times 10 = 4.5 \rightarrow 4\text{th}$ and in a fleet of 19 the 45% boat is

$45/100 \times 19 = 8.55 \rightarrow 9\text{th}$

In the Example Yacht Race shown in Table 1 the 45% Boat is the 4$^{\text{th}}$ boat *Delta* and the *SCT* of the race is *Delta's CT* = 5684 sec = 1:34:44 (nearest second)

Median Boat

The *SCT* is the *median*[4] of the *CTs* sorted in ascending order from least to greatest. The median, unlike the mean or average, is not skewed by a small proportion of extremely large or small values and is the value separating the lower half from the higher half of *CTs*. In a fleet of 10 yachts (an even number), the *SCT* is the average of the *CTs* of the 5th and 6th placed yachts. In a fleet of 19 (an odd number), the *SCT* is the *CT* of the yacht finishing in 10th place.

In the Example Yacht Race shown in Table 1 where there are 10 yachts, the *SCT* of the race is the average of the *CTs* of *Echo* and *Foxtrot* (5$^{\text{th}}$ and 6$^{\text{th}}$ yachts), equal to 5698 sec = 1:34:58 (nearest second).

The three methods discussed above; Trimmed Fleet Average, 45% Boat and Median Boat are representative of various methods used in performance handicap systems that tend to 'find' a standard boat – and hence a Standard Corrected Time – that is close to the middle of a fleet on corrected time order. And, consequently, roughly half of the fleet who finish before the standard boat will have their handicaps increased for the next race and the other half, finishing after, will have their handicaps decreased.

Another way of determining the Standard Corrected Time *SCT*

Another way of determining the Standard Corrected Time – denoted here as ***Optimum Boat*** **–** is to use mathematical optimization, where a function of the fleet's performance indicators is minimized, leading to a method for calculating the *SCT*.

Optimum Boat

Consider a yacht race as a system where handicaps and elapsed times (*AHCs* and *ETs*) are used to produce corrected time results, and then the handicaps and elapsed times are used again to update the handicaps for the next race. This updating process consists of two parts:

(i) calculation of the Standard Corrected Time *SCT* followed by the Back-Calculated Handicaps *BCHs,* then the Performance Indicators *PIs* and,

---

[3] *TopYacht* (https://topyacht.com.au/web) founded by Rod McCubbin, Cheltenham VIC 3192, and now a division of Northstar Technologies Australia, Mount Lawley WA 6929, provides race management and scoring software to the Australian sailing community.

[4] The *median* of a sample of *n* values is obtained by first ordering the values from least to greatest and then choosing the middle value if *n* is odd or the average of the two middle values if *n* is even. In either case there will be the same number of values that are larger than or equal to the median, and smaller than or equal to the median. The median is a robust estimator of the location of a sample of values drawn from a large population.

(ii) a function of the *PIs* and *AHCs* gives the Calculated Handicaps *CHCs* that become the Allocated Handicaps *AHCs* for the next race.

Our *Optimum Boat* method of determining the *SCT* uses a mathematical optimization process known as *M-estimation*[5] to produce a simple equation for the *SCT* involving summations of *weighted* handicaps and functions of elapsed times. But, the weights, denoted w_k are functions of the *SCT,* Allocated Handicaps and Elapsed Times, therefore calculating the *SCT* is iterative and concludes when the changes to the weights from one iteration to the next become acceptably small. M-estimation is explained in some detail in Appendix B where the equations we list below are developed.

For our purposes, we define a quantity y, its estimate $\hat{y}$ (where the caret symbol '^' denotes an estimate) and a residual v as follows. For the $k = 1,2,3,\dots,n$ yachts in the race, consider their allocated handicaps to be quantities y_k and their back-calculated handicaps to be estimates $\hat{y}_k$ and write a simple equation for each yacht as

$$y_k + v_k = \hat{y}_k \tag{5}$$

where the *residual* v_k is a small random quantity and (5) can be rearranged as

$$v_k = \hat{y}_k - y_k \tag{6}$$

and we may say that the residuals are functions of *measurements* and *parameters*. For example, in our PHS yacht race, (5) represents $AHC_k + v_k = BCH_k$ and equation (6) represents $v_k = BCH_k - AHC_k$, and with (3) and (4) we may write

$$v_k = SCT\big/ET_k - AHC_k = PI_k \tag{7}$$

and the residuals are functions of the parameters *SCT* (unknown), AHC_k (known), and the measurement ET_k.

M-estimators result from optimizing an *objective function* $\varphi = \sum_{k=1}^{n} \rho\left(v_k\right)$ where $\rho\left(v\right)$ is an arbitrary function of the residuals v having certain desirable characteristics and $\rho\left(v\right)$ is connected to an *influence function* $\psi\left(v\right)$ by the relationship $\psi\left(v\right) = \frac{d}{dv}\rho\left(v\right)$. The influence function $\psi\left(v\right)$ is also connected to a *weight function* $w\left(v\right)$ by the relationship $\psi\left(v\right) = v\,w\left(v\right)$ and $0 \le w \le 1$.

In our case, we choose a weight function known as *Tukey's bisquare weight function*[6] (see Appendix B) which can be defined as

$$w_k = \begin{cases} \left(1 - \left(PI_k/\gamma\right)^2\right)^2 & \text{for} \quad \left|PI_k\right| \le \gamma \\ 0 & \text{for} \quad \left|PI_k\right| > \gamma \end{cases} \tag{8}$$

where $\gamma = c\,S$ is a *weight constant,* (the symbol γ is the Greek letter gamma). S is a measure of scale calculated from the data and c is a *tuning constant* (more about S and c in Appendix B).

---

[5] M-estimation is a robust method of determining estimates of quantities that are functions of measurements. The name is a contraction of maximum likelihood estimation.

[6] John W Tukey (1915 – 2000) was an American mathematician and statistician, best known for developing the fast Fourier Transform (FFT) algorithm and box plot. The Tukey range test, the Tukey lambda distribution, and the Tukey test of additivity all bear his name. He is also credited with coining the term *bit* and the first published use of the word *software*.

$c = 4.685$ and $S = 1.4826(\text{MAD})$ where MAD is the *Median Absolute Deviation* of the performance indicators (see Appendix B) and

$$\text{MAD} = \text{median}\{|PI_k - M|\} \quad \text{where} \quad M = \text{median}\{PI_k\} \tag{9}$$

where $|x|$ means the absolute value of x and the braces $\{\ \}$ indicate a finite sample of n values. Hence, the weight constant $\gamma = cS = 6.946(\text{MAD})$

After choosing the weight function, we use the relationships between the functions $\rho(v)$, $\psi(v)$, and $w(v)$ to derive the equation for $\rho(v_k)$ as

$$\rho(v_k) = \frac{\gamma^2}{6}\begin{cases} 1 - \left(1 - \left(PI_k/\gamma\right)^2\right)^3 & \text{for } |PI_k| \le \gamma \\ 1 & \text{for } |PI_k| > \gamma \end{cases} \tag{10}$$

Now the objective function φ is

$$\varphi = \sum_{k=1}^{n} \rho(v_k) = \frac{\gamma^2}{6}\sum_{k=1}^{n}\left\{\begin{matrix} 1 - \left(1 - \left(PI_k/\gamma\right)^2\right)^3 & \text{for } |PI_k| \le \gamma \\ 1 & \text{for } |PI_k| > \gamma \end{matrix}\right\} \tag{11}$$

and since $PI_k = SCT/ET_k - AHC_k$ then the objective function φ is a function of the single unknown parameter SCT, i.e., $\varphi = \varphi(SCT)$ and φ will have an optimum (either a minimum or a maximum value) when the derivative of φ with respect to the SCT is equal to zero, i.e., $\dfrac{d\varphi}{d\,SCT} = 0$.

This leads to the solution for the standard corrected time SCT as

$$SCT = \sum \frac{w_k\, AHC_k}{ET_k} \Bigg/ \sum \frac{w_k}{ET_k^2} \tag{12}$$

where the symbol Σ (Greek letter capital sigma) represents summations that may be shown in equivalent forms as $\sum x_k = \sum_k x_k = \sum_{k=1}^{n} x_k = x_1 + x_2 + x_3 + \cdots + x_n$

For numerical stability, let $\dfrac{ET_k}{ET_{AVE}} = q_k$ where $ET_{AVE} = \dfrac{1}{n}\sum ET_k$ then

$$SCT = ET_{AVE} \sum \frac{w_k\, AHC_k}{q_k} \Bigg/ \sum \frac{w_k}{q_k^2} \tag{13}$$

The curve of the objective function φ

The objective function φ given by (11) can be expressed in a different form by using the weight function (8) and assuming that all $|PI_k| \le \gamma$ then $\varphi = \sum_{k=1}^{n} \rho(v_k) = \dfrac{\gamma^2}{6}\sum_{k=1}^{n}\left\{1 - w_k\left(1 - \left(PI_k/\gamma\right)^2\right)\right\}$ and since $PI_k = SCT/ET_k - AHC_k$ then, with a bit of algebra, we may write

$$\varphi = \frac{\gamma^2}{6}\sum_{k=1}^{n}(1 - w_k) + \frac{1}{6}\sum_{k=1}^{n}\left\{\frac{w_k}{ET_k^2} SCT^2 - 2\frac{w_k\, AHC_k}{ET_k} SCT + w_k\, AHC_k^2\right\}$$

Gathering terms gives

$$\varphi=\left[\frac{1}{6}\sum_{k=1}^{n}\frac{w_k}{ET_k^2}\right]SCT^2-\left[\frac{1}{3}\sum_{k=1}^{n}\frac{w_k\,AHC_k}{ET_k}\right]SCT+\left[\frac{\gamma^2}{6}\sum_{k=1}^{n}\left(1-w_k\right)+\frac{1}{6}\sum_{k=1}^{n}w_k\,AHC_k^2\right] \tag{14}$$

Hence, φ is an equation of the 2nd degree in the parameter SCT. This is demonstrated in Figure 1 where the curve is defined by plotted values of φ (y-axis) for 100 assumed *SCTs* (x-axis) for the yachts in the Example Yacht Race in Table 1. The *SCTs* are evenly spaced between the corrected times of the first and last yachts.

Note that the optimum value of φ is given by evaluating $\dfrac{d\varphi}{d\,SCT}=0$ and from (14) this gives

$\dfrac{d\varphi}{d\,SCT}=\left[\dfrac{1}{3}\displaystyle\sum_{k=1}^{n}\dfrac{w_k}{ET_k^2}\right]SCT-\left[\dfrac{1}{3}\displaystyle\sum_{k=1}^{n}\dfrac{w_k\,AHC_k}{ET_k}\right]=0$ and by rearrangement we obtain

$SCT=\displaystyle\sum_{k=1}^{n}\dfrac{w_k\,AHC_k}{ET_k}\Bigg/\sum_{k=1}^{n}\dfrac{w_k}{ET_k^2}$ that is equation (12)

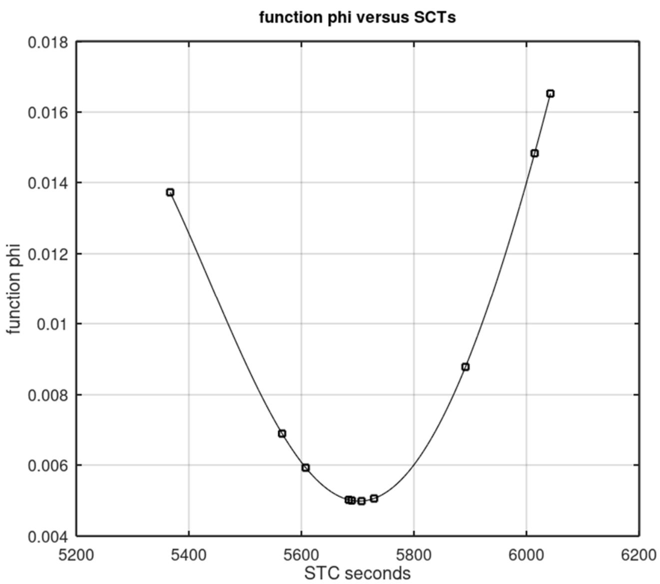

Figure 1. Curve of φ versus SCT for the Example Yacht Race in Table 1. Squares indicate an SCT equal to the corrected time of a yacht and the optimum value of $\varphi=0.0050$ is the minimum for a $SCT=5705.898$ seconds.

In M-estimation, the weights w_k are functions of the 'unknown' parameters which means that the solution is iterative and usually begins by assuming some initial values for either the weights or the parameters and then calculating a new set of weights and parameters and the iterative process ceases when differences between successive solutions reach acceptably small values.

The method of calculating the *Optimum Boat* Standard Corrected Time

1. Calculate the average of the Elapsed Times of the $k = 1,2,3,\ldots,n$ boats that finished the race

$$ET_{AVE} = \frac{1}{n}\sum_{k=1}^{n} ET_k \tag{15}$$

2. Calculate the ratios q_k where

$$q_k = ET_k \big/ ET_{AVE} \tag{16}$$

3. Set the iteration counter $iter = 1$
4. Set a small value for a quantity ε (Greek letter epsilon) for testing when the iterative process will cease. For our purposes $\varepsilon = 0.001$

Begin the iterative process.

5. **IF** iteration counter $iter = 1$, **THEN** set 'previous' weights $w_k^{(prev)} = 1$, **ELSE** $w_k^{(prev)} = w_k^{(next)}$.
6. Calculate the Standard Corrected Time SCT using

$$SCT = ET_{AVE} \sum \frac{w_k^{(prev)} AHC_k}{q_k} \Bigg/ \sum \frac{w_k^{(prev)}}{q_k^2} \tag{17}$$

7. Calculate the Back-Calculated Handicaps BCH_k using (3)

$$BCH_k = SCT \big/ ET_k \tag{18}$$

8. Calculate the Performance Indicators PI_k using (4)

$$PI_k = BCH_k - AHC_k \tag{19}$$

9. Calculate the *median* M of the Performance Indicators,

$$M = \text{median}\left(PI_k\right) \tag{20}$$

10. Calculate the *Median Absolute Deviation* MAD of the Performance Indicators where

$$\text{MAD} = \text{median}\left(\left|PI_k - M\right|\right) \tag{21}$$

11. Calculate the *weighting constant* γ (gamma) where

$$\gamma = 6.946\left(\text{MAD}\right) \tag{22}$$

12. Calculate *weights* for the next iteration $w_k^{(next)}$ where

$$w_k^{(next)} = \begin{cases} \left(1 - \left(PI_k / \gamma\right)^2\right)^2 & \text{for } \left|PI_k\right| \le \gamma \\ 0 & \text{for } \left|PI_k\right| > \gamma \end{cases} \tag{23}$$

13. **IF** all $\left|w_k^{(next)} - w_k^{(prev)}\right| < \varepsilon$, **THEN** go to step 14, **ELSE** $iter = iter + 1$, go to step 5
14. The iterative process has concluded and SCT for iteration $iter$ is the Standard Corrected Time for the yacht race.

Using the Example Yacht Race shown in Table 1, the average ET and the q values (steps 1 and 2 above) are shown in Table 2A. These are followed by the calculated values in Table 2B for the first iteration of the process outlined above (steps 3 to 13).

Sail No.	Yacht	Elapsed Time ET (sec)	Allocated Handicap AHC	Corrected Time CT (sec)	Place	q
10	*Juliet*	4997	1.074	5366.778	**1**	0.868062
7	*Golf*	5549	1.003	5565.647	**2**	0.963954
1	*Alfa*	5197	1.079	5607.563	**3**	0.902806
4	*Delta*	5639	1.008	5684.112	**4**	0.979588
5	*Echo*	5661	1.005	5689.305	**5**	0.983410
6	*Foxtrot*	5684	1.004	5706.736	**6**	0.987406
9	*India*	5834	0.982	5728.988	**7**	1.013463
2	*Bravo*	6156	0.957	5891.292	**8**	1.069400
8	*Hotel*	6344	0.948	6014.112	**9**	1.102059
3	*Charlie*	6504	0.929	6042.216	**10**	1.129853
		average 5756.500				

Table 2A. The average ET and q-values where $ET_{AVE} = 5756.5$ sec and $q_k = ET_k / ET_{AVE}$

Sail No.	prev weight $w_k^{(prev)}$	w/q	$(w/q)AHC$	$(w/q)/q$	BCH	PI	abs deviations $\lvert PI - M \rvert$	next weight $w_k^{(next)}$
10	1	1.151991	1.237239	1.327084	1.141	0.067	0.066335	0.614129
7	1	1.037394	1.040506	1.076187	1.027	0.024	0.023861	0.943712
1	1	1.107658	1.195163	1.226907	1.097	0.018	0.017437	0.969396
4	1	1.020837	1.029004	1.042108	1.011	0.003	0.002466	0.999216
5	1	1.016870	1.021954	1.034024	1.007	0.002	0.001537	0.999645
6	1	1.012755	1.016806	1.025673	1.003	-0.001	0.001537	0.999868
9	1	0.986716	0.968716	0.973608	0.977	-0.005	0.005321	0.997622
2	1	0.935104	0.894894	0.874419	0.926	-0.031	0.031428	0.908439
8	1	0.907393	0.860208	0.823362	0.899	-0.049	0.049868	0.776014
3	1	0.885071	0.822231	0.783350	0.876	-0.053	0.052971	0.749115
			sum 10.086961	sum 10.186722		M 0.00037	MAD 0.020649	

Table 2B. The first iteration of the solution for the SCT where

$$SCT = ET_{AVE} \sum \frac{w_k^{(prev)} AHC_k}{q_k} \Bigg/ \sum \frac{w_k^{(prev)}}{q_k^2} = 5756.50 \times 10.086961 / 10.186722 = 5700.125 \text{ sec},$$

$$\gamma = 6.946(\text{MAD}) = 6.946 \times 0.020649 = 0.143428,$$

$$w_k^{(next)} = \left(1 - \left(PI_k / \gamma\right)^2\right)^2 \text{ since all } \lvert PI_k \rvert \le \gamma .$$

Four iterations are sufficient to give the *Optimum Boat SCT* = 5706 sec = 1:35:06 (nearest second) and the *SCTs* and weights $w_k^{(prev)}$ and $w_k^{(next)}$ for each iteration are shown in Table 3.

	Iteration 1	Iteration 2	Iteration 3	Iteration 4
	$w_k^{(prev)}$	$w_k^{(prev)}$	$w_k^{(prev)}$	$w_k^{(prev)}$
Juliet	1.000	0.6141290553856193	0.6062991894282684	0.6046353496641509
Golf	1.000	0.9437123037963773	0.9401273207101736	0.9393542812085032
Alfa	1.000	0.9693956163322419	0.9664596806986721	0.9658208641116867
Delta	1.000	0.999216127976095	0.9987081876236733	0.9985843884508177
Echo	1.000	0.9996448425283547	0.9992828086086153	0.9991899519181419
Foxtrot	1.000	0.9998684820368631	0.9999877723767444	0.9999967312569082
India	1.000	0.9976216624695033	0.9983252306151149	0.9984585101836251
Bravo	1.000	0.9084387483165871	0.9130632549794919	0.9140287284755588
Hotel	1.000	0.7760140751165618	0.7830250864350125	0.7844991748253646
Charlie	1.000	0.7491148190258168	0.7563451199065648	0.7578667294795469
	SCT = 5700.125 sec	*SCT* = 5704.716 sec	*SCT* = 5705.691 sec	*SCT* = 5705.898 sec
	$w_k^{(next)}$	$w_k^{(next)}$	$w_k^{(next)}$	$w_k^{(next)}$
	0.6141290553856193	0.6062991894282684	0.6046353496641509	0.604282950670038 5
	0.9437123037963773	0.9401273207101736	0.9393542812085032	0.9391900499574978
	0.9693956163322419	0.9664596806986721	0.9658208641116867	0.9656848960208297
	0.999216127976095	0.9987081876236733	0.9985843884508177	0.9985574588312575
	0.9996448425283547	0.9992828086086153	0.9991899519181419	0.9991695730903568
	0.9998684820368631	0.9999877723767444	0.9999967312569082	0.999997896059553
	0.9976216624695033	0.9983252306151149	0.9984585101836251	0.9984860132391258
	0.9084387483165871	0.9130632549794919	0.9140287284755588	0.9142324444264633
	0.7760140751165618	0.7830250864350125	0.7844991748253646	0.7848106807522991
	0.7491148190258168	0.7563451199065648	0.7578667294795469	0.7581883415900925

Table 3. *SCTs* and weights $w_k^{(prev)}$ and $w_k^{(next)}$ for each iteration

Because of the iterative nature of the solution for the Standard Corrected Time *SCT,* a computer language supporting user-defined functions would be an appropriate choice. Examples of two options are shown below, one is a function written in GNU Octave[7] called **`yacht_M_estimate.m`** and the other is a Microsoft Excel macro called **`MestimateSCT()`** written in Visual Basic for Applications (VBA).

Octave function to compute SCT

Because of the iterative nature of the solution for the *SCT,* a computer language supporting user-defined functions would be an appropriate choice. One such language is GNU Octave[8] and a function called **`yacht_M_estimate.m`** is shown below. It requires a list of Elapsed Times in seconds (ET_sec) and Allocated Handicaps *AHCs* for the n yachts in the race and returns the *SCT,* weights, a flag (M_flag = 1 for success and M_flag = 0 for failure to converge), and the number of iterations required for an acceptable solution.

The Octave and Matlab languages support matrix algebra where a matrix is a two-dimensional (2D) data structure consisting of rows and columns. Vectors are matrices with either a single column (a column vector) or a single row (a row vector). The words matrix and array are often used interchangeably, but an array is a more general data structure containing elements of a single type in one, two, three, or higher dimensions.

[7] **GNU Octave** is a high-level language, primarily intended for numerical computations. It provides a convenient command line interface for solving linear and nonlinear problems numerically, and for performing other numerical experiments using a language that is mostly compatible with **Matlab**. GNU Octave is freely redistributable software from the Free Software Foundation.

[8] **GNU Octave** is a high-level language, primarily intended for numerical computations. It provides a convenient command line interface for solving linear and nonlinear problems numerically, and for performing other numerical experiments using a language that is mostly compatible with **Matlab**. GNU Octave is freely redistributable software from the Free Software Foundation.

Arithmetic operations of addition, subtraction and multiplication of matrices follow rules of matrix algebra, but division requires matrix inversion which is a more complicated process. In the function `yacht_M_estimate.m` shown below the matrices are vectors (1D arrays) and arithmetic operations of multiplication and division with vectors of the same length are simpler and achieved with element-by-element operations. For example, vector `x` divided by a constant `a` is shown as `x./a` and vector `x` divided by the elements of vector y is `x./y`, and similarly for vector multiplication.

```
function [SCT,w,M_flag,iter] = yacht_M_estimate(ET_sec,AHC)
%
% This function computes the Standard Corrected Time (SCT) of a yacht race
% using M-estimation with Tukey's bisquare weight function
%
%-------------------------------------------------------------------------------
% Function:  yacht_M_estimate
%
% Usage:     [SCT w M_flag,iter] = yacht_M_estimate(ET_sec,AHC);
%
% Author:    Rod Deakin,
%            13 Aldercress Approach,
%            DUNSBOROUGH, WA 6281, AUSTRALIA.
%            email: randm.deakin@gmail.com
%            Version  1.0   16 June 2023
%                     1.1   04 December 2024
%
% Functions required:
%    none
%
% Purpose:
%   This function computes the Standard Corrected Time (SCT) of a yacht race
%   given a set of elapsed times (ET) in seconds and allocated handicaps (AHC)
%
% Variables:
%    AHC         - vector of Allocated Handicaps
%    b           - multiplier for MAD (b = 1.4826)
%    BCH         - vector of Back Calculated Handicaps
%    c           - tuning constant (c = 4.685)
%    epsilon     - small value for testing difference between weights
%    ET_sec      - vector of elapsed times in seconds
%    ET_ave      - average of elapsed times (seconds)
%    gamma       - weight constant (gamma = c*S)
%    iter        - iteration number
%    k           - integer counter
%    M           - median of PIs
%    MAD         - median absolute deviation
%    M_flag      - 1 = Pass, 0 = Fail
%    n           - number of yachts in race
%    PI          - vector of Performance Indicators where PI = BCH - AHC
%    q           - vector of ratios ET/ET_ave
%    S           - scale of PIs (S = b*MAD)
%    SCT         - Standard Corrected Time (seconds)
%    u           - standardised PI (u = PI/gamma)
%    w           - vector of weights
%    w_prev      - vector of previous weights
%    w_next      - vector of next weights
%
% Remarks:
%    The function returns SCT, w, M_flag, and iter.
%    weights are in the range  0 <= w < 1
%    If M_flag = 1 then the iterative procedure has converged to a valid result
%                  for SCT
%    If M_flag = 0 then the iterative procedure has failed to converge in 20
%                  iterations to a valid result for SCT
%
% References:
```

```
%    [1] Deakin, R.E., (2024), 'Another way to Calculate the Standard Corrected
%           Time of a yacht Race', xx pages, 04-Dec-2024.  Available at URL
%           http://www.mygeodesy.id.au in the Sailing folder
%
%--------------------------------------------------------------------------------

% determine number of yachts in race
n = length(AHC);
% calculate the average of the elapsed times (in seconds)
ET_ave = mean(ET_sec);
% calculate the ratios q = ET/ET_ave
q = ET_sec./ET_ave;
% set the maximum number of iterations
max_iter = 20;
% set tuning constant
c = 4.685;
% set multiplier for MAD where S = b*MAD
b = 1.4826;
% set iteration counter
iter = 1;
%set value of epsilon
epsilon  = 0.001;
% set vector of weights at unity for initial solution
w_prev = ones(n,1);

while (1)

  % determine the SCT
  x = w_prev./q;
  sum1 = sum(x.*AHC);
  sum2 = sum(x./q);
  SCT  = ET_ave*sum1/sum2;

  % compute BCHs and PIs
  BCH = SCT./ET_sec;
  PI = BCH-AHC;

  % compute median of performance indicators
  M = median(PI);
  % compute Median Absolute Deviation
  MAD = median(abs(PI-M));
  % compute weighting factor gamma
  S = b*MAD;
  gamma = c*S;

  % compute new weights w_next using Tukey's bisquare weight function
  w_next = zeros(n,1);
  for k = 1:n
    u = PI(k)/gamma;
    if (abs(u) > 1)
      w_next(k) = 0;
    else
      w_next(k) = (1-u^2)^2;
    endif
  end

  % test the new weights w_next to see if iterative process has converged
  if abs(w_next-w_prev) < epsilon
    M_flag = 1;
    w = w_prev;
    break;
  endif
  if iter > max_iter
    fprintf('\nIteration for weights failed to converge after %2d iterations\n\n',max_iter);
    M_flag = 0;
    w = w_prev;
    break;
  endif
```

```
  % update weights and iteration number
  w_prev = w_next;
  iter = iter+1;
endwhile
endfunction
```

Excel Macro to compute SCT

Visual Basic for Applications (VBA) is an implementation of Microsoft's object-oriented programming language Visual Basic 6.0 built into most desktop Microsoft Office applications. Using VBA in Excel allows the incorporation of user-defined functions in Excel processes. The VBA program below is in the form of a **sub**, also known as a subroutine or sub procedure, and the code is used to perform a specific task but does not return any value. In Excel, this code is referred to as a Macro.

```
Sub MestimateSCT()
'
'-------------------------------------------------------------------------------
' This Excel subroutine, written in Microsoft's Visual Basic for Applications,
' computes the Standard Corrected Time (SCT) of a yacht race using M-estimation
' with Tukey's bisquare weight function
'
' Subroutine:  MestimateSCT
'
' Author:      Rod Deakin,
'              13 Aldercress Approach,
'              DUNSBOROUGH, WA 6281, AUSTRALIA.
'              email: randm.deakin@gmail.com
'              Version  1.0  04 December 2024
'
' Purpose:
'   This function computes the Standard Corrected Time (SCT) of a yacht race
'   given a set of elapsed times (ET) and allocated handicaps (AHC)
'
' Variables:
'    AHC         - Allocated Handicap
'    ET          - elapsed time
'    aveET       - average of elapsed times
'    Iter        - iteration counter (integer)
'    Gamma       - weight constant
'    j           - integer counter
'    M           - median of PInd's
'    MAD         - median absolute deviation
'    n           - number of yachts in race
'    PInd        - Performance Indicator where PInd = SCT/ET - AHC
'    q           - vector of ratios ET/aveET
'    Row         - row counter (integer)
'    SCT         - Standard Corrected Time (seconds)
'    Test        - sum of values that are either 1 or 0 for each yacht.
'                  When Test = n the iterative sequence has converged
'    u           - standardised PInd (u = PInd/Gamma)
'    Weight      - weight
'    prevWeight  - previous weight
'    nextWeight  - next weight
'
' Remarks:
'    The subroutine requires an Excel workbook with at least two spreadsheets,
'    one named "Race Times" and the other named "Optimum Boat".
'    The spreadsheet named Race Times must have the following form
'    * 6 columns A,B,C,D,E,F that are: (A)Sail No., (B)Yacht Name,
'      (C)Elapsed Time, (D)Allocated Handicap, (E)Corrected Time, (F)Place
'    * Rows 1 & 2 contain race series information
'    * Rows 3 & 4 are column headers
'    * Rows 5 and onwards contain the race data
'    * Elapsed Times are assumed to be in Excel Custom format h:mm:ss
'    * Corrected Times are assumed to be in Excel Custom format hh:mm:ss.000
'
```

```
'    An example of a "Race Times" spreadsheet for n = 10 yachts is shown below
'
'        |     A     |     B  |        C     |    D      |      E         |   F   |
'       1| World Sailing  - International Empirical Handicap Scheme for Yachts
'       2|
'       3|           |        | Elapsed Time | Allocated | Corrected Time |       |
'       4| Sail No. |  Yacht |  (h:mm:ss)   | Handicap  |    (h:mm:ss)   | Place |
'       5|    10        J       1:23:17       1.074      01:29:26.778      1
'       6|     7        G       1:32:29       1.003      01:32:45.647      2
'       7|     1        A       1:26:37       1.079      01:33:27.563      3
'       8|     4        D       1:33:59       1.008      01:34:44.112      4
'       9|     5        E       1:34:21       1.005      01:34:49.305      5
'      10|     6        F       1:34:44       1.004      01:35:06.736      6
'      11|     9        I       1:37:14       0.982      01:35:28.988      7
'      12|     2        B       1:42:36       0.957      01:38:11.292      8
'      13|     8        H       1:45:44       0.948      01:40:14.112      9
'      14|     3        C       1:48:24       0.929      01:40:42.216     10
'
'    Weights are in the range  0 <= W < 1 and are calculated using Tukey's
'    formula for the Biweight where
'       Weight = (1-(PInd/Gamma)^2)^2 if |PInd| <= Gamma
'       Weight = 0 if |PInd| > Gamma
'
'    Since PInd = ET/SCT - AHC then both PInd and Weight are functions of
'    the unknown SCT and the process of solving for the SCT is iterative and
'    converges to an acceptable solution when the change in weights from one
'    iteration to the next is less than 0.001.
'
'    The subroutine copies the data from sheet "Race Times" to sheet
'    "Optimum Boat" and intermediate calculations are shown in columns after
'    the Place column on this sheet.  At the conclusion of the iterative
'    process the SCT is written below the column of Corrected Times.
'    The median M of the PInd's and the Median Absolute Deviation (MAD) where
'    MAD = Median(|PInd-M|), are shown and the value of Gamma is written below
'    the MAD value.  In the column after nextWeight are values that
'    are either 1 or 0. And if |prevWeight-nextWeight|<0.001 than value = 1, or
'    0 otherwise.  At the bottom of this column is the sum of these values
'    (that equals n when the process has converged) and below that is the number
'    of iterations for convergence.
'    The value for the SCT is also written below the Corrected Times in the
'    'Race Times" spreadsheet.
'
' References:
'    [1] Deakin, R.E., (2024), 'Another way to Calculate the Standard Corrected
'           Time of a yacht Race', xx pages, 04-Dec-2024.  Available at URL
'           http://www.mygeodesy.id.au in the Sailing folder
'
'--------------------------------------------------------------------------------
'
'Select the Race Times sheet.
    Sheets("Race Times").Select

'Determine the row number of the last yacht in the race
    J = 4
    For Row = 5 To 250
      If (Sheets("Race Times").Cells(Row, 1).Value <> "") Then
        J = J + 1
      End If
    Next Row

'Calculate the number of yachts in the race
    n = J - 4

'Copy the column headers and race data from sheet "Race Times"
'to sheet "Optimum Boat"
    Sheets("Race Times").Select
    Range(Cells(3, 1), Cells(J, 6)).Select
    Selection.Copy
```

```
      Sheets("Optimum Boat").Select
      Range("A1").Select
      ActiveSheet.Paste
   'Use AutoFit to adjust the column width
      Sheets("Optimum Boat").Select
      Range(Cells(1, 1), Cells((n + 2), 6)).Select
      Selection.Columns.AutoFit
      Range("A1").Select

  'Calculate the average of the Elapsed Times
      Range(Cells(3, 3), Cells((n + 2), 3)).NumberFormat = "hh:mm:ss"
      aveET = WorksheetFunction.Average(Range(Cells(3, 3), Cells((n + 2), 3)))
      Cells((n + 4), 3).NumberFormat = "hh:mm:ss.000"
      Cells((n + 4), 3).Value = aveET

  'Calculate the ratios q = ET / aveET
      Cells(2, 7).HorizontalAlignment = xlCenter
      Cells(2, 7).VerticalAlignment = xlCenter
      Cells(2, 7).Value = "q"
      For Row = 3 To (n + 2)
        ET = Cells(Row, 3).Value
        Cells(Row, 7).Value = ET / aveET
      Next Row

  'Iteration routine to determine SCT
      Test = 0
      Iter = 0
      Do While Test < n
        Iter = Iter + 1
        'Set Weights
        Cells(2, 8).HorizontalAlignment = xlCenter
        Cells(2, 8).VerticalAlignment = xlCenter
        Cells(2, 8).Value = "prevWeight"
        If Iter = 1 Then  'Set prevWeight = 1
          For Row = 3 To (n + 2)
            Cells(Row, 8).Value = 1
          Next Row
        Else  'prevWeight = nextWeight
          For Row = 3 To (n + 2)
            Cells(Row, 8).Value = Cells(Row, 11).Value
          Next Row
        End If
        'Calculate SCT
        Sum1 = 0
        Sum2 = 0
        For Row = 3 To (n + 2)
          q = Cells(Row, 7).Value
          Weight = Cells(Row, 8).Value
          AHC = Cells(Row, 4).Value
          Sum1 = Sum1 + (Weight * AHC / q)
          Sum2 = Sum2 + (Weight / (q * q))
        Next Row
        SCT = aveET * Sum1 / Sum2
        Cells((n + 4), 8).NumberFormat = "hh:mm:ss.000"
        Cells((n + 4), 8).Value = SCT

        'Calculate Performance Indicators (PInd)
        Cells(2, 9).HorizontalAlignment = xlCenter
        Cells(2, 9).VerticalAlignment = xlCenter
        Cells(2, 9).Value = "PInd"
        For Row = 3 To (n + 2)
          ET = Cells(Row, 3)
          AHC = Cells(Row, 4)
          PInd = (SCT / ET) - AHC
          Cells(Row, 9).Value = PInd
        Next Row
        'Calculate the median of the PInd's
        M = WorksheetFunction.Median(Range(Cells(3, 9), Cells((n + 2), 9)))
```

```
        Cells((n + 4), 9).Value = M
        'Calculate the absolute deviations |PInd - M|
        Cells(2, 10).HorizontalAlignment = xlCenter
        Cells(2, 10).VerticalAlignment = xlCenter
        Cells(2, 10).Value = "|PInd - M|"
        For Row = 3 To (n + 2)
          Cells(Row, 10).Value = Abs(Cells(Row, 9) - M)
        Next Row
        'Calculate MAD = median of absolute deviations |PInd - M|
        MAD = WorksheetFunction.Median(Range(Cells(3, 10), Cells((n + 2), 10)))
        Cells((n + 4), 10).Value = MAD
        'Calculate the weighting constant Gamma
        Gamma = 6.946 * MAD
        Cells((n + 5), 10).Value = Gamma
        'Calculate nextWeight
        Cells(2, 11).HorizontalAlignment = xlCenter
        Cells(2, 11).VerticalAlignment = xlCenter
        Cells(2, 11).Value = "nextWeight"
        For Row = 3 To (n + 2)
          PInd = Cells(Row, 9).Value
          If Abs(PInd) > Gamma Then
            Weight = 0
          Else
            u = PInd / Gamma
            u2 = u * u
            Weight = (1 - u2) ^ 2
          End If
          Cells(Row, 11).Value = Weight
        Next Row

        'Test for convergence of weights

        'Calculate the absolute value of the weight difference |prevWeight - nextWeight|
        Cells(2, 12).HorizontalAlignment = xlCenter
        Cells(2, 12).VerticalAlignment = xlCenter
        Cells(2, 12).Value = "|DiffWeight|"
        Cells(2, 13).HorizontalAlignment = xlCenter
        Cells(2, 13).VerticalAlignment = xlCenter
        Cells(2, 13).Value = "Test Flag"
        For Row = 3 To (n + 2)
          prevWeight = Cells(Row, 8).Value
          nextWeight = Cells(Row, 11).Value
          absDiffWeight = Abs(prevWeight - nextWeight)
          Cells(Row, 12).Value = absDiffWeight
          If absDiffWeight < 0.001 Then
            Cells(Row, 13).Value = 1
          Else
            Cells(Row, 13).Value = 0
          End If
        Next Row
        Test = WorksheetFunction.Sum(Range(Cells(3, 13), Cells((n + 2), 13)))
        Cells((n + 4), 13).Value = Test

    Loop

    Cells((n + 5), 13).Value = Iter

'Select the Race Times sheet and print the SCT at the bottom of the Corrected Times
    Sheets("Race Times").Select
    Cells((n + 6), 4).HorizontalAlignment = xlRight
    Cells((n + 6), 4).VerticalAlignment = xlCenter
    Cells((n + 6), 4).Value = "SCT = "
    Cells((n + 6), 5).NumberFormat = "hh:mm:ss.000"
    Cells((n + 6), 5).Value = SCT

End Sub
```

Comparison of Standard Corrected Times

Using the Example Yacht Race (Table 1) we could have the following Standard Corrected Times *SCTs*

		Standard Boat	
Method	Standard Corrected Time	Real	Imaginary
Trimmed Fleet Average	5672 sec = 1:34:32 (h:mm:ss)		(3)-(4)
45% Boat	5684 sec = 1:34:44	*Delta* (4)	
Median Boat	5698 sec = 1:34:58		(5)-(6)
Optimum Boat	5706 sec = 1:35:06		(5)-(6)

Table 4. Standard Corrected Times for the Example Yacht Race.
[For the Trimmed Fleet Average, the standard boat is imaginary with a corrected time between the 3rd and 4th place finishers. The standard boat for the 45% Boat is Delta, which was 4th on corrected time.]

Discussion

The computation of the Standard Corrected Time *SCT* of a yacht race using the proposed *Optimum Boat* method is not simple but achievable with programming languages, e.g. Octave functions and VBA Excel macros. The technique (and its solution) assumes an *objective function* φ is a function of the k^{th} yacht's performance indicator PI_k, its elapsed time ET_k, its allocated handicap AHC_k, and a *weight* w_k that is a function of the yacht's *PI*. But a yacht's *PI* is a function of the unknown *SCT* which means φ is a function of unknown weights w_k and a single unknown parameter *SCT*. The optimum value of φ is its minimum value and this is when $SCT = ET_{AVE} \sum \frac{w_k \, AHC_k}{q_k} \Big/ \sum \frac{w_k}{q_k^2}$ and the solution for the *SCT* is iterative since w_k is also a function of the *SCT*. An algorithm for the iterative solution is provided and a by-product of the solution is a set of weights $\{w_1 \quad w_2 \quad w_3 \quad \cdots \quad w_n\}$ that are numbers in the range $[0,1]$.

Small or zero weights are associated with larger-than-normal *PIs,* which indicate anomalous *AHCs* or perhaps incorrect *ETs*. Thus, the weights can be used to assess the quality of the allocated handicaps, identify incorrect ones, and *indicate possible spurious ETs.*

For example, Table 5 shows a portion of the output from the Excel macro MestimateSCT() for the Example Yacht Race in Table 1.

Sail No.	**Yacht**	**Elapsed Time**	**Allocated Handicap**	**Corrected Time**	**Place**	**q**	**prevWeight**
10	*Juliet*	01:23:17	1.074	01:29:26.778	1	0.868062	0.6046
7	*Golf*	01:32:29	1.003	01:32:45.647	2	0.963954	0.9394
1	*Alfa*	01:26:37	1.079	01:33:27.563	3	0.902806	0.9658
4	*Delta*	01:33:59	1.008	01:34:44.112	4	0.979588	0.9986
5	*Echo*	01:34:21	1.005	01:34:49.305	5	0.983410	0.9992
6	*Foxtrot*	01:34:44	1.004	01:35:06.736	6	0.987406	1.0000
9	*India*	01:37:14	0.982	01:35:28.988	7	1.013463	0.9985
2	*Bravo*	01:42:36	0.957	01:38:11.292	8	1.069400	0.9140
8	*Hotel*	01:45:44	0.948	01:40:14.112	9	1.102059	0.7845
3	*Charlie*	01:48:24	0.929	01:40:42.216	10	1.129853	0.7579

aveET = 01:35:56.500 SCT = 01:35:05.897

Table 5. Optimum Boat data for Example Yacht Race

All times in Table 5 are in h:mm:ss format. The average elapsed time $ET_{AVE} = $ 1:35:56.500 and column **q** contains ratios $q_k = ET_k / ET_{AVE}$. The column headed **prevWeights** are the weights for the beginning of the 4th iteration (see Table 3) and the Standard Corrected Time $SCT = $ 1:35:05.897 .

Now suppose the *AHC* of *Juliet* was incorrectly entered as 1.704 giving the output as shown in Table 6

Sail No.	**Yacht**	**Elapsed Time**	**Allocated Handicap**	**Corrected Time**	**Place**	**q**	**prevWeight**
10	*Juliet*	01:23:17	**1.704**	01:29:26.778	1	0.868062	**0.0000**
7	*Golf*	01:32:29	1.003	01:32:45.647	2	0.963954	0.9294
1	*Alfa*	01:26:37	1.079	01:33:27.563	3	0.902806	0.9523
4	*Delta*	01:33:59	1.008	01:34:44.112	4	0.979588	0.9919
5	*Echo*	01:34:21	1.005	01:34:49.305	5	0.983410	0.9932
6	*Foxtrot*	01:34:44	1.004	01:35:06.736	6	0.987406	0.9968
9	*India*	01:37:14	0.982	01:35:28.988	7	1.013463	0.9994
2	*Bravo*	01:42:36	0.957	01:38:11.292	8	1.069400	0.9623
8	*Hotel*	01:45:44	0.948	01:40:14.112	9	1.102059	0.8817
3	*Charlie*	01:48:24	0.929	01:40:42.216	10	1.129853	0.8634

aveET = 01:35:56.500 SCT = 01:35:45.552

Table 6. Incorrect *AHC* for *Juliet*

The weight of *Juliet* is zero which could be an indicator of either an incorrect *ET* or incorrect *AHC.*

Conclusion

A Performance Handicap System (PHS) uses a set of rules and mathematical calculations to produce a series of Corrected Times *CTs* for a yacht race, the Standard Corrected Time *STC* for that race, and Allocated Handicaps *AHCs* for the next race. The *STC* can be established according to arbitrary rules, e.g., *World Sailing's Trimmed Fleet Average,* or *TopYacht's 45% Boat,* or by a mathematical rule, e.g., the *Median Boat.*

A new method, *Optimum Boat* has been presented in this paper. It relies on finding a minimum value of a function of the race fleet's Performance Indicators *PIs* that yields the *SCT* and a set of weights that can be used in 'data snooping'. This method does not use arbitrary rules, instead, it relies only on the fleet's *PIs.* The solution of the *SCT* and the weights, which is iterative, has been set out in the paper with examples of computer functions provided.

Optimum Boat could be an attractive alternative to the usual (and often arbitrary) rules for establishing the *SCT* of a yacht race.

References

Banas, M. and Ligas, M., (2014), 'Empirical tests of performance of some M-estimators', *Geodesy and Cartography,* Vol. 63, No. 2, pp. 127-146.
https://journals.pan.pl/Content/98355/PDF/art1_Banas_ligas.pdf

Beaton, A.E. and Tukey, J.W., (1974), 'The fitting of power series, meaning of polynomials, illustrated on band-spectroscopic data', *Technometrics,* Vol. 16, No. 2 (May, 1974), pp. 147-185.

Deakin, R.E. and Green, R.A., 2023, 'Notes on Performance Handicap Systems in Yachting' 49 pages, 22-Feb-2023.
http://www.mygeodesy.id.au in the Sailing folder [accessed 04-Dec-2024]

Hogg, R.V., (1979), 'Statistical robustness: One view of its use in applications today', *The American Statistician,* Vol. 33, No. 3, pp. 108-115.
https://www.soa.org/globalassets/assets/library/research/actuarial-research-clearing-house/1978-89/1979/arch-3/arch79v34.pdf [accessed 04-Dec-2024]

Huber, P.J., (1964), 'Robust estimation of a location parameter', *Annals of Mathematical Statistics,* Vol. 35, pp. 73-111.

——————— (1981), *Robust Statistics,* John Wiley & Sons, New York.

TopYacht, 2021, 'How the Next Handicap is Calculated', *TopYacht* Technical Documents and Discussion Papers, 15 pages, 01-Apr-2021.
https://topyacht.net.au/results/shared/technical/How%20the%20Next%20Handicap%20is%20Calculated.pdf?ty=5679 [accessed 04-Dec-2024]

World Sailing, 2016, 'International Empirical Handicap Scheme for Yachts', *World Sailing,* 3 pages.
https://www.sailing.org/tools/documents/TurnkeytextVer2-[7780].pdf [accessed 04-Dec-2024]

Note that the Excel spreadsheet calculations in Example 2 can be found at URL
https://www.sailing.org/tools/documents/EHWebRaceResultandNumberAdjustmentVers2-[20432].xlsx

(This Excel workbook contains 2 spreadsheets, **Points** and **Race 6**. The workbook was created in 2013 and last modified on 15-Mar-2016. The author is Ken Kershaw)

Yohai, V.J., (1987), 'High breakdown-point and high efficiency robust estimates for regression', *The Annals of Statistics,* Vol. 15, No. 20, pp. 642-656.
https://projecteuclid.org/download/pdf_1/euclid.aos/1176350366

APPENDIX A

WORLD SAILING[9] – INTERNATIONAL EMPIRICAL HANDICAP SCHEME FOR YACHTS

Welcome to the World Sailing Empirical Handicap Scheme for Yachts. As the name suggests the scheme is intended to permit yachts, generally displacement boats with keels, of varying designs to race against each other and after racing determine, by calculation, the race results by excluding the performance differences of the boats themselves. The scheme is an empirical handicap scheme, that is a scheme where after racing the relative performance of each boat - their handicap, is determined from the times it took each boat to complete the race.

World Sailing provides this scheme to any race organiser who wishes to use it. It is intended to operate in isolation at local/race organiser level requiring no input to or from World Sailing or elsewhere. World Sailing does however offer users a basic method of handicap allocation to a boat for use in its first race.

Before using the scheme an organiser needs to address four factors:-

- The allocation of a boats handicap for its first race
- How to calculate race results
- How to adjust a boats handicap after racing
- Whether or not to attempt to exclude the varying skills of crews from the calculations

The allocation of a boats handicap for its first race

A boats handicap is expressed as a number based about 1. Faster boats handicaps will be higher than 1 with slow boats handicaps less than 1. Generally, the range of handicaps will be no more than 1.2 and no less than 0.8.

It would never be wrong for a race organiser to allocate a first race handicap based on their own subjective opinion of a boat. If the organiser considers the boat to be of average performance, then a handicap of 1 would suit. If, however the organiser considers the boat faster than the fleet average then a handicap above 1 in the range of say 1 to1.2 would be appropriate. If the performance is thought to be below average, then a handicap of less than 1 in the range of 0.8 to 1 should be used.

Alternatively, if the race organiser wishes the first race handicap could be allocated using the basic calculator at the following link - .

Whatever the case the handicap number used to calculate the race results for a boat in its first and subsequent races should be adjusted before use in the boats next race.

How to calculate race results

The results of a race are determined by comparing the *corrected times* for each boat with the least time being the race winner, the next least second place and so on for each boat completing in the race.

The *corrected time* (CT) for each boat is calculated by multiplying its elapsed time (ET), that is the time it took to complete the race, by its handicap (H) i.e. CT = ET x H

[9] *World Sailing* is the governing body for the sport of sailing formed in 1907 and then known as the International Yacht Racing Union (IYRU). The name was changed to the International Sailing Federation (ISAF) in 1996 before adopting the name *World Sailing* in 2015.

An example of the calculations and how best to set this out is shown below.

Example 1 – Race Results

Sail No	Boat	Finish Time	Elapsed Time (ET)	Handicap (H)	Corrected Time (CT)	Finishing Place
1	A	14:56:37	01:26:37	1.079	01:33:28	3
2	B	15:12:36	01:42:36	0.957	01:38:11	8
3	C	15:18:24	01:48:24	0.929	01:40:42	9
4	D	15:03:59	01:33:59	1.008	01:34:44	4
5	E	15:04:21	01:34:21	1.005	01:34:49	5
6	F	15:04:44	01:34:44	1.004	01:35:07	6
7	G	15:02:29	01:32:29	1.003	01:32:46	2
8	H	15:15:44	01:45:44	0.948	01:40:14	10
9	I	15:07:14	01:37:14	0.982	01:35:29	7
10	J	14:53:17	01:23:17	1.074	01:29:27	1

Start Time = 13:30:00

How to adjust a boats handicap after racing

The life blood of empirical handicap racing is the adjustment of handicaps after racing. Without this race results and the scheme will soon become meaningless.

The World Sailing empirical handicap scheme attempts to adjust the handicap of each boat based on the *standard corrected time* (SCT) of the fleet which is the average CT excluding the lowest 20% and highest 40% of the CTs (rounded down to whole numbers).

Using the race result example above those CTs exclude are flagged in blue and yellow as shown below. The remaining CTs are averaged to give a SCT for the race (1:34:32 in the example).

Dividing the SCT by a boats ET gives the calculated handicap which the boat would have had in the race for its CT to have equaled the SCT i.e. it gives the handicap to which the boat sailed in the race (h).

The difference between H and h gives a performed indicator (PI) i.e. PI = h - H (which may be plus or minus). A proportion of the PI should be applied to the boats race handicap (H) with the result used as the boats new handicap in its next race (H’).

The portion of the PI applied to adjust the handicap depends on the number of races the boat has completed in the fleet. The table below gives the portions. The new handicap H’ = H + (PI x PM).

Races completed	Portion	Multiplier
1	All	1
2	Half	0.5
3	One third	0.33
4	One quarter	0.25
5	One fifth	0.2
Greater than 5	One fifth	0.2

Example 2 – Race Results and Number Adjustment

Sail No	Boat	Finish Time f	Elapsed Time ET	Handicap H	Corrected Time CT	Finishing Place	CTs used for SCT	Performed Handicap h	Performed Indicator PI	PI Multiplier PM	New Handicap H'
input	input	input	f - ST	input	ET x H	input	input	SCT / ET	h - H	input	H + (PI x PM)
1	A	14:56:37	1:26:37	1.079	1:33:28	3	1:33:28	1.091	0.012	0.2	1.081
2	B	15:12:36	1:42:36	0.957	1:38:11	8		0.921	-0.036	0.2	0.950
3	C	15:18:24	1:48:24	0.929	1:40:42	9		0.872	-0.057	1	0.872
4	D	15:03:59	1:33:59	1.008	1:34:44	4	1:34:44	1.006	-0.002	0.2	1.008
5	E	15:04:21	1:34:21	1.005	1:34:49	5	1:34:49	1.002	-0.003	0.25	1.004
6	F	15:04:44	1:34:44	1.004	1:35:07	6	1:35:07	0.998	-0.006	0.25	1.002
7	G	15:02:29	1:32:29	1.003	1:32:46	2		1.022	0.019	0.33	1.009
8	H	15:15:44	1:45:44	0.948	1:40:14	10		0.894	-0.054	0.5	0.921
9	I	15:07:14	1:37:14	0.982	1:35:29	7		0.972	-0.010	0.5	0.977
10	J	14:53:17	1:23:17	1.074	1:29:27	1		1.135	0.061	0.2	1.086

Start Time (ST) = 1:30:00 PM

SCT = 1:34:32

Whether or not to attempt to exclude the varying skills of crews from the calculations

Unlike a Rating System an Empirical Handicap Scheme of the type explained here allocates handicaps to the combined boat performance and the crew skill. This can sometimes work to the detriment of good crews and benefit of not so good crews as their ability is reflected in the adjusted handicaps. Whether or not to attempt to exclude crew skill from the calculations is a decision for the race organiser bearing in mind that to attempt this mathematically will involve on-going subjective judgements on the part of the organiser. For more information on the exclusion of crew skill from the calculations please contact World Sailing at – technical@sailing.org

This document at URL https://www.sailing.org/tools/documents/TurnkeytextVer2-[7780].pdf (accessed 04-Dec-2024).

Note that the Excel spreadsheet calculations in Example 2 can be found at URL https://www.sailing.org/tools/documents/EHWebRaceResultandNumberAdjustmentVers2-[20432].xlsx (This Excel workbook contains 2 spreadsheets, **Points** and **Race 6**. The workbook was created in 2013 and last modified on 15-Mar-2016. The author is Ken Kershaw)

APPENDIX B

M-Estimation

An estimator is a rule (a set of equations perhaps) for calculating an estimate of a quantity from observed data. An estimator is efficient if its estimates are calculated in some 'best possible' manner, and it is unbiased if the difference between the expected value of the estimate and its true value is zero. For our purposes, we define a quantity y, its estimate $\hat{y}$ (where the caret symbol '^' denotes an estimate) and a residual v as follows.

For the $k = 1,2,3,\dots,n$ yachts in the race, consider their allocated handicaps to be quantities y_k and their back-calculated handicaps be estimates $\hat{y}_k$ and write a simple equation for each yacht as

$$y_k + v_k = \hat{y}_k \tag{24}$$

where the *residual* v_k is a small random quantity and (24) can be rearranged as

$$v_k = \hat{y}_k - y_k \tag{25}$$

and we may say that the residuals are functions of *measurements* and *parameters*.

For example, in our Performance Handicap System yacht race, equation (24) represents $AHC_k + v_k = BCH_k$ and equation (25) represents $v_k = BCH_k - AHC_k$, and with (3) and (4) we may write

$$v_k = SCT/ET_k - AHC_k = PI_k \tag{26}$$

and the residuals are functions of the single parameter *SCT*, and the measurements ET_k and AHC_k.

M-Estimators, originally proposed by Huber (1964) are a group of estimators that are outcomes of optimizing *objective functions* φ having the general form

$$\varphi = \sum_{k=1}^{n} \rho\left(v_k\right) \tag{27}$$

where $\rho\left(v_k\right)$ is an arbitrary function of the residuals v_k having certain desirable characteristics, and a reasonable $\rho\left(v_k\right)$ should have the following properties.

- Always non-negative, $\rho\left(v_k\right) \geq 0$
- Equal to zero when its argument is zero, $\rho\left(0\right) = 0$
- Symmetric, $\rho\left(v_k\right) = \rho\left(-v_k\right)$
- Monotone in $\left|v_k\right|$ for $0 < \left|v_k\right| < \left|v_{k+1}\right| \;\Rightarrow\; \rho\left(\left|v_k\right|\right) \leq \rho\left(\left|v_{k+1}\right|\right)$
- Differentiable

In M-estimation, a *weight* w is obtained from a function of the residuals v and is a numeric value representing the degree of importance attached to a quantity, greater values reflecting more importance, and $0 \leq w \leq 1$. The *weight function* is defined as

$$w\left(v\right) = \psi\left(v\right)/v \tag{28}$$

$\psi\left(v\right)$ is the *influence function* that is the derivative of $\rho\left(v\right)$ or

$$\psi\left(v\right) = \frac{d}{dv}\rho\left(v\right) \tag{29}$$

The inter-relationship between the three functions $\left(\rho\text{-}, \psi\text{-}, w\text{-functions}\right)$ would allow the ρ-function to be determined from the w-function by first determining the ψ-function from (28) as $\psi\left(v\right) = v\,w\left(v\right)$ and then the ρ-function from (29) as $\rho\left(v\right) = \int \psi\left(v\right)dv$. Alternatively, the ψ-function could be defined and then $\psi\left(v\right) = v\,w\left(v\right)$ and $w\left(v\right) = \psi\left(v\right)\big/v$.

We have chosen to use a weighting function, commonly known as *Tukey's bisquare weight function* or *biweight* introduced by Beaton and Tukey (1974) and defined as

$$w\left(u\right) = \begin{cases} \left(1 - u^2\right)^2 & \text{for} \quad \left|u\right| \le 1 \\ 0 & \text{for} \quad \left|u\right| > 1 \end{cases} \tag{30}$$

where u is a scaled residual defined as

$$u_k = v_k\big/\gamma = \left(\hat{y}_k - y_k\right)\big/\left(c\,S\right) \tag{31}$$

and $\gamma = c\,S$ is a *weight constant,* and the symbol γ is the Greek letter gamma. S is a measure of the scale that is calculated from the data and c is a *tuning constant* (more about S and c later). Using (31) we may write the weighting function (30) as

$$w\left(v\right) = \begin{cases} \left(1 - \left(v\big/\gamma\right)^2\right)^2 & \text{for} \quad \left|v\right| \le \gamma \\ 0 & \text{for} \quad \left|v\right| > \gamma \end{cases} \tag{32}$$

Now using (28) gives $\psi\left(v\right) = v\,w\left(v\right)$ and with (32) we may write

$$\psi\left(v\right) = \begin{cases} v\left(1 - \left(v\big/\gamma\right)^2\right)^2 & \text{for} \quad \left|v\right| \le \gamma \\ 0 & \text{for} \quad \left|v\right| > \gamma \end{cases} \tag{33}$$

Integrating both sides of (29) gives $\rho\left(v\right) = \int \psi\left(v\right)dv$ and with (33) $\rho\left(v\right) = \int v\left(1 - \left(v\big/\gamma\right)^2\right)^2 dv$.

To evaluate this integral let $s = 1 - \left(v\big/\gamma\right)^2$ then $ds = -\left(2v\right)\big/\gamma^2\,dv$ and $vdv = -\gamma^2\big/2\,ds$ so that $\rho\left(v\right) = -\gamma^2\big/2\int s^2 ds = -\left(\gamma^2\big/2\right)\left(s^3\big/3\right) + C = -\dfrac{\gamma^2}{6}\left(1 - \left(v\big/\gamma\right)^2\right)^3 + C$ where C is a constant of integration. From our previous discussion, a desirable property of $\rho\left(v\right)$ is that $\rho\left(0\right) = 0$ and enforcing this condition means that the constant $C = \gamma^2\big/6$ and

$$\rho\left(v\right) = \frac{\gamma^2}{6}\begin{cases} 1 - \left(1 - \left(v\big/\gamma\right)^2\right)^3 & \text{for} \quad \left|v\right| \le \gamma \\ 1 & \text{for} \quad \left|v\right| > \gamma \end{cases} \tag{34}$$

[Note that we can confirm (29) if we differentiate (34) with respect to v.

Make the substitution $s = 1 - \left(v\big/\gamma\right)^2$ then $ds\big/dv = -2v\big/\gamma^2$ and $\rho\left(v\right) = \dfrac{\gamma^2}{6}\left(1 - s^3\right)$.

Now $d\rho\big/dv = d\rho\big/ds \cdot ds\big/dv = \left(-\gamma^2\big/2 \cdot s^2\right)\left(-2v\big/\gamma^2\right) = vs^2$ and $\dfrac{d}{dv}\rho\left(v\right) = v\left(1 - \left(v\big/\gamma\right)^2\right)^2$ as expected.]

The M-estimator (in this case a single equation) will arise from optimizing the objective function φ, i.e.,

$$\varphi = \sum_{k=1}^{n} \rho\left(v_k\right) \Rightarrow \text{optimum}$$

And using (34) the objective function φ is

$$\varphi = \sum_{k=1}^{n} \rho\left(v_k\right) = \frac{\gamma^2}{6} \sum_{k=1}^{n} \left\{ \begin{array}{ll} 1 - \left(1 - \left(v_k/\gamma\right)^2\right)^3 & \text{for } \left|v\right| \le \gamma \\ 1 & \text{for } \left|v\right| > \gamma \end{array} \right\} \tag{35}$$

On the assumption that all $\left|v_k\right| \le \gamma$ and that $v_k = SCT/ET_k - AHC_k$ we write

$$\begin{aligned} \varphi = \sum_{k=1}^{n} \rho\left(v_k\right) &= \frac{\gamma^2}{6} \sum_{k=1}^{n} \left\{ 1 - \left(1 - \left(v_k/\gamma\right)^2\right)^3 \right\} \\ &= \frac{\gamma^2}{6} \sum_{k=1}^{n} \left\{ 3\left(v_k/\gamma\right)^2 - 3\left(v_k/\gamma\right)^4 + \left(v_k/\gamma\right)^6 \right\} \\ &= \frac{\gamma^2}{6} \sum_{k=1}^{n} \left\{ \frac{3}{\gamma^2}\left(SCT/ET_k - AHC_k\right)^2 - \frac{3}{\gamma^4}\left(SCT/ET_k - AHC_k\right)^4 + \frac{1}{\gamma^6}\left(SCT/ET_k - AHC_k\right)^6 \right\} \end{aligned}$$

Now φ is a function of the estimated parameter *SCT*, i.e., $\varphi = \varphi\left(SCT\right)$ and the function will be an optimum (either a minimum or a maximum value) when the derivative with respect to the *SCT* is equal to zero, i.e., $\varphi \Rightarrow \text{optimum}$ when $\dfrac{d\varphi}{d\,SCT} = 0$ and the derivative is obtained as follows

$$\begin{aligned} \frac{d\varphi}{d\,SCT} &= \frac{\gamma^2}{6} \sum_{k=1}^{n} \left\{ \begin{array}{l} \frac{6}{\gamma^2}\left(SCT/ET_k - AHC_k\right)\left(\frac{1}{ET_k}\right) - \frac{12}{\gamma^4}\left(SCT/ET_k - AHC_k\right)^3\left(\frac{1}{ET_k}\right) \\ + \frac{6}{\gamma^6}\left(SCT/ET_k - AHC_k\right)^5\left(\frac{1}{ET_k}\right) \end{array} \right\} \\ &= \sum_{k=1}^{n} \frac{1}{ET_k} \left\{ \left(SCT/ET_k - AHC_k\right) - \frac{2}{\gamma^2}\left(SCT/ET_k - AHC_k\right)^3 + \frac{1}{\gamma^4}\left(SCT/ET_k - AHC_k\right)^5 \right\} \\ &= \sum_{k=1}^{n} \frac{1}{ET_k} \left(SCT/ET_k - AHC_k\right) \left\{ 1 - \frac{2}{\gamma^2}\left(SCT/ET_k - AHC_k\right)^2 + \frac{1}{\gamma^4}\left(SCT/ET_k - AHC_k\right)^4 \right\} \\ &= \sum_{k=1}^{n} \frac{1}{ET_k} \left(SCT/ET_k - AHC_k\right) \left(1 - \frac{1}{\gamma^2}\left(SCT/ET_k - AHC_k\right)^2\right)^2 \end{aligned} \tag{36}$$

But, from (26) and (32) we have the weighting function

$$w_k = \left\{ \begin{array}{lll} \left(1 - \frac{1}{\gamma^2}\left(SCT/ET_k - AHC_k\right)^2\right)^2 = \left(1 - \left(PI_k/\gamma\right)^2\right)^2 & \text{for} & \left|PI_k\right| \le \gamma \\ 0 & \text{for} & \left|PI_k\right| > \gamma \end{array} \right. \tag{37}$$

And substituting (37) into (36) gives

$$\frac{d\varphi}{d\,SCT} = \sum_{k=1}^{n} \frac{w_k}{ET_k} \left(SCT/ET_k - AHC_k\right) \tag{38}$$

The weights w_k are positive numeric values less than or equal to 1, or they are zero, and even though they are functions of the *SCT* they can be considered as constants for any particular value of the *SCT* and the second derivative becomes $\dfrac{d^2\varphi}{d\,SCT^2} = \sum_{k=1}^{n} \dfrac{w_k}{ET_k^2} > 0$ and the optimum value of φ will be a minimum when the derivative (38) is equated to zero. This leads to the equation for SCT as

$$SCT = \sum \frac{w_k\,AHC_k}{ET_k} \Bigg/ \sum \frac{w_k}{ET_k^2} \tag{39}$$

For numerical stability, let $\dfrac{ET_k}{ET_{AVE}} = q_k$ where $ET_{AVE} = \dfrac{1}{n}\sum ET_k$ then

$$SCT = ET_{AVE} \sum \frac{w_k\,AHC_k}{q_k} \Bigg/ \sum \frac{w_k}{q_k^2} \tag{40}$$

In M-estimation, the weights w_k are functions of the 'unknown' parameters which means that the solution is iterative and usually begins by assuming some initial values for either the weights or the parameters and then calculating a new set of weights and parameters and the iterative process ceases when differences between successive solutions reach acceptably small values.

In our case, there is a single unknown parameter, the *SCT,* and the iterative process begins by assuming a set of initial weights all equal to unity and ceases when the changes to the weights reach acceptably small values.

Tukey's bisquare weight function with Median Absolute Deviation (MAD)

In the calculation of the weights w_k using Tukey's bisquare weight function (37) the factor $\gamma = cS$ is required where c is *tuning constant* (see below) and S is a measure of scale (or variability of the data) computed from the data. For a sample of size n, measures of scale S of the residuals v_k could be the sample standard deviation s_v computed from the sample variance $s_v^2 = \dfrac{1}{n-1}\sum_{k=1}^{n}\left(v_k - \overline{v}\right)^2$ where s_v is the positive square root of s_v^2 and $\overline{v} = \dfrac{1}{n}\sum_{k=1}^{n} v_k$ is the sample mean which is a measure of location (or centre) of the sample. But the sample mean and variance (and hence the sample standard deviation) are known to suffer from the effects of outliers, since large residuals affect the mean $\overline{v}$ and also the squared differences $\left(v_k - \overline{v}\right)^2$ in the calculation of the variance.

A more robust measure of the location of a sample is the *median M* and a more robust measure of the scale is the *Median Absolute Deviation* (MAD) which is defined as the median of the absolute deviations from the sample's median *M,* i.e.,

$$\text{MAD} = \text{median}\left\{\left|v_k - M\right|\right\} \quad \text{where } M = \text{median}\left\{v_k\right\} \tag{41}$$

where the braces $\{\ \}$ indicate a finite sample of n values.

The median M of a sample $\left\{x_j\right\}$ of n values ordered from smallest to largest so that $x_1 < x_2 < \ldots < x_n$ is

$$M\left\{x_j\right\} = \begin{cases} x_{p+1} & \text{if } n = 2p+1 \text{ is odd} \\ \frac{1}{2}\left(x_p + x_{p+1}\right) & \text{if } n = 2p \text{ is is even} \end{cases} \tag{42}$$

In either case, there will be the same number values that are larger than or equal to the median, and smaller than or equal to the median *M.*

For example, suppose v_j is a set of $j = 1,2,\ldots,n$ values and for $n = 7$, $v_j = \{-2 \quad 7 \quad 4 \quad 16 \quad 1 \quad 0 \quad 8\}$.

The set is ordered from smallest to largest as $\{-2 \quad 0 \quad 1 \quad \underset{\uparrow}{4} \quad 7 \quad 8 \quad 16\}$ and since n is odd, the median M is the middle value indicated with $\uparrow$, $p = (n-1)/2 = 3$ and $M = v_{p+1} = 4$. There are 3 values less than M (the values to the left of the 4$^{\text{th}}$ value) and 3 values greater than M (the values to the right of the 4$^{\text{th}}$ value).

Now suppose the set $v_j = \{-2 \quad 7 \quad 2 \quad 16 \quad 1 \quad 0 \quad 8 \quad 4 \quad 4 \quad -5\}$ has $n = 10$ values that are ordered from smallest to largest as $\{-5 \quad -2 \quad 0 \quad 1 \quad \underset{\uparrow}{2} \quad \underset{\uparrow}{4} \quad 4 \quad 7 \quad 8 \quad 16\}$, and since n is even, the median M is the average of the two middle values and $p = n/2 = 5$ and $M = \frac{1}{2}(v_p + v_{p+1}) = 3$. There are 5 values less than the median (the first 5 values) and 5 values greater than the median (the last 5 values).

It should be noted here that if X is a random variable that can take values n values $x_1, x_2, \ldots, x_n$ having a median M then the probability that any X is less than or equal to the median is exactly $\frac{1}{2}$ or $\Pr(X \le M) = \frac{1}{2}$ and if X is a continuous random variable with a probability density function $f_X(x)$ and cumulative distribution function $F_X(x)$, so that $F_X(x) = \int_{-\infty}^{x} f_X(y)\,dy$, or $\frac{d}{dx}F_X(x) = f_X(x)$ then the median M is defined by the solution of the integral equation $\Pr(X \le M) = F_X(M) = \int_{-\infty}^{M} f_X(x)\,dx = \frac{1}{2}$.

Appendix C shows how this result can be used to determine the value of a scale factor b that enables the MAD to be used as a consistent estimator of the standard deviation σ of normally distributed data where

$$\hat{\sigma} = b \times \text{MAD} \approx 1.4826(\text{MAD}) \tag{43}$$

The measure of scale S above and in (31) is often taken to be $S = 1.4826(\text{MAD})$.

The Tuning Constant in M-estimation

M-estimation is the outcome of optimizing the objective function $\varphi = \sum_{k=1}^{n} \rho(v_k)$ where $\rho(v_k)$ is a function of the residuals v_k and is related to the influence function $\psi(v)$ and weight function $w(v)$ by $\psi(v) = \frac{d}{dv}\rho(v)$ and $w(v) = \frac{\psi(v)}{v}$. The residuals v_k are defined from the general relationship measurement + residual = best estimate (or $y_k + v_k = \hat{y}_k$) giving $v_k = \hat{y}_k - y_k$ and a scaled residual $u_k = \frac{v_k}{\gamma} = \frac{\hat{y}_k - y_k}{cS}$ where $\gamma = cS$ and S is a measure of scale computed from the residuals and c is a tuning constant.

Now suppose that the residuals are each divided by S, computed from the sample, and these *standardized* residuals are $\tilde{v}_k = \frac{v_k}{S}$ and the scaled residuals $u_k = \frac{\tilde{v}_k}{c}$. For example, using Tukey's bisquare weight function

$$w(\tilde{v}) = \begin{cases} \left(1 - \left(\tilde{v}/c\right)^2\right)^2 & \text{for } |\tilde{v}| \le c \\ 0 & \text{for } |\tilde{v}| > c \end{cases} \tag{44}$$

the ψ and ρ functions are

$$\psi\left(\tilde{v}\right)=\begin{cases}\tilde{v}\left(1-\left(\tilde{v}/c\right)^2\right)^2 & \text{for } \left|\tilde{v}\right|\le c\\ 0 & \text{for } \left|\tilde{v}\right|>c\end{cases} \tag{45}$$

$$\rho\left(\tilde{v}\right)=\frac{c^2}{6}\begin{cases}1-\left(1-\left(\tilde{v}/c\right)^2\right)^3 & \text{for } \left|\tilde{v}\right|\le c\\ 1 & \text{for } \left|\tilde{v}\right|>c\end{cases} \tag{46}$$

The M-estimator $\psi\left(\tilde{v}\right)$, resulting from optimizing $\varphi=\sum_{k=1}^{n}\rho\left(v_k\right)$, should be an unbiased estimator and its *efficiency* can be defined as a ratio of the minimum possible variance of an unbiased estimator to the actual variance of the estimator and it can be proved that this ratio is less than or equal to unity, i.e., for an unbiased estimator $\hat{\theta}$,

$$\mathit{eff}\left(\hat{\theta}\right)=\frac{\text{minimum possible variance of }\hat{\theta}}{\text{actual variance of }\hat{\theta}}\le 1$$

The actual variance of $\hat{\theta}$ can only be determined if the probability distribution of the random variable, from which the estimator is derived, is known. Hence the efficiency of an estimator is described as 'relative to' or 'with respect to' a particular distribution. The standard normal distribution is often assumed to be the underlying probability distribution.

The efficiency of an estimator is often expressed as a percentage, e.g. if $\mathit{eff}\left(\hat{\theta}\right)=0.95$ then $\hat{\theta}$ has an efficiency of 95% with respect to the standard normal distribution.

An equation for 95% efficiency of an M-estimator, assuming the residuals are from a standard normal distribution, is given by Huber (1981) as

$$\mathit{eff}=\frac{\left[\int_{-c}^{c}\psi'\left(x\right)f_X\left(x\right)dx\right]^2}{\int_{-c}^{c}\left[\psi\left(x\right)\right]^2 f_X\left(x\right)dx}\approx 0.95 \tag{47}$$

where $x\sim N\left(0,1\right)$ are the random variables, $f_X\left(x\right)$ is the pdf of the standard normal distribution, $\psi\left(x\right)$ is the influence function for any M-estimator and $\psi'\left(x\right)=\frac{d}{dx}\psi\left(x\right)$

Equation (47) involving the tuning constant c as integration limits is solved numerically by Banas & Ligas (2014) to obtain $c=4.685$ for the influence function for Tukey's biweight (see (44) to (46) above with x replacing $\tilde{v}$). For example, with

$$\psi\left(x\right)=\begin{cases}x\left(1-\left(x/c\right)^2\right)^2 & \text{for } \left|x\right|\le c\\ 0 & \text{for } \left|x\right|>c\end{cases}\quad\text{then}\quad\psi'\left(x\right)=\begin{cases}\left(1-\left(x/c\right)^2\right)\left(1-5\left(x/c\right)^2\right) & \text{for } \left|x\right|\le c\\ 0 & \text{for } \left|x\right|>c\end{cases}$$

And with $f_X\left(x\right)=\frac{1}{\sqrt{2\pi}}e^{-\frac{1}{2}x^2}$ equation (47) can be (rather crudely) evaluated using the following function **`eff`** written in GNU Octave[10]

```
function eff
for c = 4.68:0.005:4.70
  sumx = 0;
  sumy = 0;
  dx = 0.0005;
  root = sqrt(2*pi);

  for x = -c:dx:c
    fx = 1/root*exp(-x*x/2);
    u  = x/c;
    u2 = u*u;
    px = x*(1-u2)^2;
    pdashx = (1-u2)*(1-5*u2);
    sumx = sumx + (pdashx*fx*dx);
    sumy = sumy + (px^2*fx*dx);
  end
  eff = sumx^2/sumy;
  fprintf(' c   = %5.3f',c);
  fprintf('\n eff = %8.6f\n',eff);
end
endfunction
```

The results, shown in the Octave Command Window, are

```
>> eff
 c   = 4.680
 eff = 0.949793
 c   = 4.685
 eff = 0.949997
 c   = 4.690
 eff = 0.950201
 c   = 4.695
 eff = 0.950403
 c   = 4.700
 eff = 0.950605
>>
```

The computed efficiency of 0.949997 for $c=4.685$ confirms the result of Banas & Ligas (2014) and others, e.g., Hogg 1979 and Yohai 1987.

[10] GNU Octave is a high-level language, primarily intended for numerical computations. It provides a convenient command line interface for solving linear and nonlinear problems numerically, and for performing other numerical experiments using a language that is mostly compatible with Matlab. GNU Octave is freely redistributable software from the Free Software Foundation.

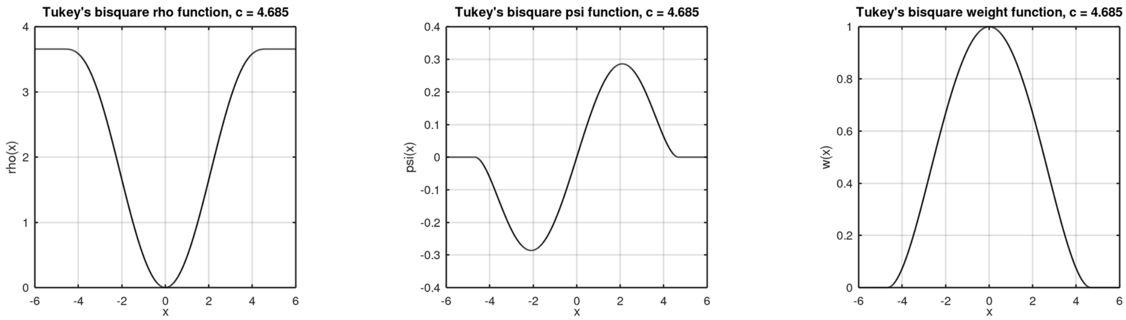

Figure 5. Tukey's bisquare ρ- , ψ- and w- functions

Appendix C

Population Median and Median Absolute Deviation (MAD)

The derivation of the probability statements $\Pr\left(\left|X-\mu\right|\le \text{MAD}\right)=\frac{1}{2}$ and $\Pr\left(Z\le\frac{\text{MAD}}{\sigma}\right)=\frac{3}{4}$ are the work of Dr Max Hunter, who turned his keen eye and talent for rigour to a topic not often treated in the statistical literature. It's a joy.

Let X be a random variable with a density function $f_X\left(x\right)$ and distribution function $F_X\left(x\right)$, so that $F_X\left(x\right)=\int\limits_{-\infty}^{x} f_X\left(y\right)dy$, or $\frac{d}{dx}F_X\left(x\right)=f_X\left(x\right)$.

The population median m is defined by the solution of the integral equation

$$\Pr\left(X\le m\right)=F_X\left(m\right)=\int\limits_{-\infty}^{m} f_X\left(x\right)dx=\frac{1}{2} \tag{48}$$

The alternative equation

$$\int\limits_{m}^{\infty} f_X\left(x\right)dx=\frac{1}{2} \tag{49}$$

can also be used to define m.

Let the random variable Y be defined by

$$0\le Y=\left|X-m\right|=\begin{cases}X-m, & \text{if } X\ge m\\ m-X, & \text{if } X<m\end{cases}$$

Suppose its density function is $g_Y\left(y\right)$ with distribution function $G_Y\left(y\right)$. Then for $y\ge 0$,

$$\begin{aligned}G_Y\left(y\right)&=\Pr\left(Y\le y\right)\\&=\Pr\left(\left|X-m\right|\le y\right)\\&=\Pr\left(-y\le X-m\le y\right)\\&=\Pr\left(m-y\le X\le m+y\right)\\&=\Pr\left(X\le m+y\right)-\Pr\left(X\le m-y\right)\\&=F_X\left(m+y\right)-F_X\left(m-y\right)\end{aligned}$$

And for $y<0$, $G_Y\left(y\right)=0$.

Hence for $y\ge 0$

$$g_Y\left(y\right)=\frac{d}{dy}G_Y\left(y\right)=\frac{d}{dy}\left\{F_X\left(m+y\right)-F_X\left(m-y\right)\right\}=f_X\left(m+y\right)+f_X\left(m-y\right)$$

and $g_Y\left(y\right)=0$ for $y<0$.

The population median M of the random variable Y satisfies the equation

$$\int\limits_{-\infty}^{M} g_Y\left(y\right)dy=\frac{1}{2}$$

And

$$\begin{aligned}\int_{-\infty}^{M} g_Y(y)\,dy &= \int_{-\infty}^{0} g_Y(y)\,dy + \int_{0}^{M} g_Y(y)\,dy = \int_{0}^{M} g_Y(y)\,dy \\ &= \int_{0}^{M} \left\{ f_X(m+y) + f_X(m-y) \right\} dy \\ &= \int_{m}^{m+M} f_X(s)\,ds - \int_{m}^{m-M} f_X(t)\,dt \qquad \text{(with substitutions } s = m+y,\ t = m-y) \\ &= \int_{m-M}^{m+M} f_X(s)\,ds \end{aligned}$$

and therefore

$$\int_{m-M}^{m+M} f_X(s)\,ds = \frac{1}{2} \tag{50}$$

Suppose now that $f_X(x)$ is symmetric about the origin then $m = 0$ from (48). So, by (50)

$$\frac{1}{2} = \int_{-M}^{M} f_X(x)\,dx = 2\int_{0}^{M} f_X(x)\,dx$$

and therefore

$$\int_{0}^{M} f_X(x)\,dx = \frac{1}{4}$$

Thus, the interval $\left[-M, M\right]$ encloses an area of 0.5 under the density function for *X*, or since

$$\int_{-\infty}^{M} f_X(x)\,dx = \frac{3}{4},$$

M is the 75 percentile of X.

But M is just the definition of MAD, so for any random variable X with a population mean $E\left\{X\right\} = \mu$ and a symmetric density function about $E\left\{X\right\} = \mu$

$$\Pr\left(\left|X-\mu\right| \le \text{MAD}\right) = \frac{1}{2} \tag{51}$$

Now

$$\begin{aligned}\Pr\left(\left|X-\mu\right| \le \text{MAD}\right) &= \Pr\left(\frac{\left|X-\mu\right|}{\sigma} \le \frac{\text{MAD}}{\sigma}\right) \\ &= \Pr\left(\left|Z\right| \le \frac{\text{MAD}}{\sigma}\right) \\ &= \Pr\left(-\frac{\text{MAD}}{\sigma} \le Z \le \frac{\text{MAD}}{\sigma}\right) \quad \text{by definition of modulus} \\ &= \Pr\left(Z \le \frac{\text{MAD}}{\sigma}\right) - \Pr\left(Z \le -\frac{\text{MAD}}{\sigma}\right) \quad \text{by symmetry} \\ &= \Pr\left(Z \le \frac{\text{MAD}}{\sigma}\right) - \left[1 - \Pr\left(Z \le \frac{\text{MAD}}{\sigma}\right)\right] \quad \text{by definition} \\ &= 2\Pr\left(Z \le \frac{\text{MAD}}{\sigma}\right) - 1\end{aligned}$$

and, using (51)

$$\Pr\left(Z \le \frac{\text{MAD}}{\sigma}\right) = \frac{3}{4} \tag{52}$$

If $f_Z\left(z\right)$ is the density function of the standard normal distribution and $F_Z\left(z\right)$ is the distribution function (see Appendix B) then

$$\frac{\text{MAD}}{\sigma} = F_Z^{-1}\left(\frac{3}{4}\right) \tag{53}$$

Where F_Z^{-1} denotes the standard normal inverse cumulative distribution function. Most mathematical software packages (Maple, Mathematica, Matlab, R, etc.) have functions to compute inverse cumulative distribution functions and for the standard normal distribution GNU Octave has a function `norminv()` that computes the value of $F_Z^{-1}\left(x\right)$ and for $\text{MAD}/\sigma = F_Z^{-1}\left(3/4\right)$ can be computed from the following instructions in the Octave Command Window.

```
>> format long g
>> MAD_on_sigma = norminv(3/4)
MAD_on_sigma = 0.6744897501960818
>>
```

And

$$\frac{\text{MAD}}{\sigma} = F_Z^{-1}\left(\frac{3}{4}\right) \approx 0.6744897501960818 \tag{54}$$

Inspection of (43) leads to

$$b = \frac{\sigma}{\text{MAD}} \approx 1.482602218505602 \tag{55}$$

We can use this relationship to estimate the standard deviation from

$$\hat{\sigma} = b \times \text{MAD} \approx 1.4826\left(\text{MAD}\right) \tag{56}$$